The following table gives for the warmer half of the year the relative frequencies of the intervals for each 10 days between 10 and 50 days:

Table 6.

	10-19	20-29	31-40	41-50
	Per cent.	*Per cent.*	*Per cent.*	*Per cent.*
1840-1879	16	26	34	23
1880-1915	17	26	34	23

The two halves of the period of 75 years show practically identical distribution. It is clearly obvious that a markedly unsymmetrical distribution has persisted throughout the entire period in summer, 43 per cent of the total number of intervals between 10 and 50 being below 30 and 57 per cent above 30. The mode for the entire period is approximately 34.

The following table gives the relative frequencies for the winter months:

Table 7.

	10-19	20-29	31-40	41-50
1840-1879	19	29	32	20
1880-1915	16	34	33	16

Thus for the entire period the distribution is practically symmetrical, with the mode slightly above 30.

The following table gives the interval frequency for the dates of maxima and minima separately for the two halves of the year:

Table 8.

	10-19	20-29	31-40	41-50	10-29	31-50
MAXIMA.						
Winter	19	28	33	20	47	53
Summer	17	27	32	24	44	56
MINIMA.						
Winter	16	35	32	17	51	49
Summer	16	24	37	22	40	59

Examination of these figures shows that in the warmer half of the year the distribution of the intervals for both maxima and minima is of marked asymmetry. In winter, however, the intervals based on the dates of minima are of practically symmetrical distribution, while the maxima yield a distribution with a slight tendency to asymmetry, not, however, so pronounced as in summer.

The difference between the results for winter and summer may be plausibly accounted for by the well-known tendency for HIGHS and LOWS to be of more intense development and rapid movement in winter. The extreme pressures are confined to a much smaller area when the systems are intensely developed and consequently a more nearly fortuitous occurrence of the dates of extreme pressures at any one locality would result. Low pressure areas, being more variable in their departures from the normal than high areas and with the extreme reading more localized, there would naturally result a greater tendency to fortuity in the dates of occurrence of lowest pressure.

Suppose, for example, in addition to the Toronto data we had similar data for Rochester. We should expect to find, as in fact we actually do, more agreement between the dates at the two places in summer than in winter and in winter there would be closer agreement between the dates of maxima than of minima.

Thus if there is a tendency for systematic recurrence at intervals somewhat greater than 30 days, as seems to be indicated by the results for the warmer season, this tendency would be modified or even entirely obliterated by the greater tendency to fortuitous occurrence in winter, particularly for extremes of low pressure.

A further compilation was made of the intervals between the dates of maximum pressure in each month and that of the second month following. The most frequent interval, if the dates were of purely fortuitous occurrence would be 60. Actually the most frequent interval was around 65, which is a double 32 to 33-day interval.

It should be understood that the results by this method are not to be interpreted as indicating the probable length of the monthly periodicity with any degree of accuracy. All that may be reasonably deduced from the facts here presented is that there is a systematic tendency for the recurrence of periods of high pressure, particularly in summer, at intervals somewhat greater than 30 days. The tendency for a purely fortuitous occurrence of the dates in winter, particularly so for the dates of minima, is what we should, *a priori*, expect. This being the case, a marked departure from a symmetrical distribution, which occurs in summer and has persisted for 75 years as shown by the close agreement of the results for the two halves of the period, can not be explained other than as a result of a systematic tendency for the dates of extremes of pressure to depart from a purely fortuitous occurrence. The question as to the actual average length of the period and to what extent it may vary in length from time to time is left unanswered. Other evidence, however, indicates that this periodicity may have a variable length over a range of a week or more and hence investigators who have observed recurrences which they regarded as of solar or lunar origin, may have been misled by the apparent coincidence of a minimum length of the monthly periodicity with solar or lunar periods. When the period resumed its normal length the apparent coincidence disappeared. Thus Koeppen's results which require for their explanation the hypothesis of a systematic tendency to a monthly periodicity are plausibly explained by variations in the length of the period.

551.501

THE MEAN VARIABILITY AS A STATISTICAL COEFFICIENT.

The difficulty of applying the ordinary Theory of Errors to meteorological computations, on account of the peculiar nature of the meteorological variables as contrasted with that of the mathematical variables,[1] has often been recognized.[2] If the arithmetic mean of a series of values is to be the value most worthy of confidence, and is to have any significance and correspond to something physical, then the individual values from which it is computed must be distributed about it according to the Law of Gauss—the deviations from the mean must obey the laws of fortuitous errors.[3]

There are two equivalent tests which are ordinarily applied in order to determine whether or not the individual deviations from the mean are due to fortuitous

[1] L. Besson: On the comparison of meteorological data with results of chance. Mo. WEATHER REV., Feb., 1920, 48: 89.

[2] V. H. Ryd: On computation of meteorological observations, *Danske Meteorologiske Institut*, 1917.

[3] Angot: *Annales du Bur. Cent. Mét.*, 1895 and 1900; and *Annuaire de la Soc. Mét.*, 51, 1903.

causes: (*a*) If there are N numbers, there should be found $f(x)N$ of which the absolute deviation is equal to or greater than x; theory gives the value of $f(x)$; (*b*) the value of the expression $\frac{2N\Sigma\delta^2}{(\Sigma\delta)^2}$ should be 3.14159 . . . Now meteorological data may satisfy both these tests without at all fulfilling other conditions equally demanded by theory; we have here a good illustration of the oft-repeated warning against drawing conclusions from summary coefficients alone, such as the mean. In the present instance, the *order* in which the numbers appear is of great significance, and the following relation must also hold:[4]

If the deviations from the mean are to be likened to fortuitous errors, then the ratio of the mean variability to the mean deviation must be equal[5] *to* $\sqrt{2}=1.414$ . . . The variabilities and deviations are taken without regard to sign.

Drawings from a sack containing balls, on each of which was marked an observed daily temperature, would give a succession vastly different from the succession actually observed: Long series of increasing or decreasing values would be less frequent in the drawing than in the observing, and the mean variability would be greater in the former; in fact the ratio of mean variability to mean deviation in the case of series of daily temperatures turns out to be but little more than half the theoretical value; chance would give the deviations which are observed, but would not give the succession which is observed. *Yet both the actual and the chance successions satisfy the two tests mentioned above.*

It has been pointed out by Besson (*op. cit.*) that if a variable is taking on random values, it does not follow that the succession of the signs of the variations will obey the laws of chance; Goutereau points out further that the deviations from the mean may not be fortuitous even if they follow the Law of Gauss.—*Edgar W. Woolard.*

[4] Ch. Goutereau: Sur la variabilité de la température, *Annuaire de la Soc. Mét. de France*, 54, 122–127, 1906.

[5] The demonstration, by Maillet, is given by Goutereau, *op. cit.* The absolute difference between a number and the next consecutive number is the variability.

551.54 : 551.501

THE VARIATE-DIFFERENCE CORRELATION METHOD.

For correlating daily changes of barometric height at Halifax and Wilmington, Miss Cave[1] made use of a formula, devised by Pearson, giving the correlation coefficient between the differences of successive daily readings at the two stations; and remarked that this formula would apply to any case in which it was desired to correlate the difference of one pair of quantities with the difference of another pair; no comments on where this procedure might be desirable were offered, however. Later, Hooker[2] independently pointed out that the correlation coefficient between two variables, for each of which a series of observations is available, is a test of similarity of the two phenomena as influenced by the totality of the causes affecting each of them; when, therefore, the observations extend over a considerable period of time, certain difficulties arise which find no precise parallel in the case where the whole of the observations refer to the same moment of time: If a diagram be drawn, showing by curves the changes of the two variables during the period under consideration, some relation will often be suggested between the usually smaller and more rapid alterations while at the same time the slower "secular" changes may or may not exhibit any similarity. If, then, the correlation coefficient be formed in the ordinary way, employing deviations from the mean, a high value will be obtained if the "secular" changes are similar (this value being almost independent of the similarity or dissimilarity of the more rapid changes), but a value approximating to zero if the "secular" changes are of quite dissimilar character even though the similarity of the smaller rapid changes be extremely marked; deductions drawn from ordinary correlation coefficients may be very erroneous. In order to get rid of the spurious correlation arising from the fact that both variables are functions of the time, the correlation coefficient may be formed between the *variations*, or first differences, of the quantities, instead of between the quantities themselves. After this method had been in rather extensive use for some time, Pearson pointed out that it was valid only when the connection between the variables and the time was linear.

The name Variate-Difference Correlation was given by Pearson[3] to a generalization of the preceding artifice, in which it was demonstrated[4] that if the variables are randomly distributed in time and space, the correlation between the variables and that between the corresponding *nth* differences will be the same; and that when this is not the case, we can eliminate variability which is due to position in time or space, and so determine whether there really is any correlation between the variables themselves, by correlating the 1st, 2d, 3d, * * *, *nth* differences: *when the correlations between the differences remain steady for several successive orders of differences we may reasonably suppose we have reached the true correlation between the variables.*

The complete theory of the method was worked out by Anderson[5] and subjected to critical examination by Pearson (*op. cit.*), who found that, as usual, the theoretical formulæ were only roughly approximated to in practice unless a great number of observations were at hand.

There has been no source more fruitful of fallacious statistical argument than the common influence of the time factor. The difference method of correlation is one of great promise and usefulness. The very frequent and superficial statements that such and such variables, both changing rapidly with the time, are essentially causative cease to have any foundation when the difference method is applied.[6]—*Edgar W. Woolard.*

[1] F. E. Cave-Browne-Cave; On the influence of the time factor on the correlation between the barometric heights at stations more than 1,000 miles apart, *Proc. Roy. Soc.*, 74:403–413, 1904–1905.

[2] R. H. Hooker: On the correlations of successive observations, *Jour. Roy. Statistical Society*, 68:696–703, 1905.

[3] Beatrice M. Cave and Karl Pearson: Numerical illustrations of the variate difference correlation method, *Biometrika*, 10, 340–355, 1914–15.

[4] "Student": The elimination of spurious correlation due to position in time or space, *Biometrika*, 10, 179–180, 1914–15.

[5] Nochmals über "The elimination of spurious correlation due to position in time or space," O. Anderson, *Biometrika*, 10, 269–279, 1914–15.

[6] Illustrations of the method are given by Cave and Pearson, *op. cit.*, and by G. U. Yule, *Introduction to the Theory of Statistics*, 5 ed., 1919, pp. 197–201; see also T. Okada, Some researches in the far eastern seasonal correlations, MO. WEATHER REV., 1917, 45: 238, 299, 535.

NOTE ON PROF. MARVIN'S DISCUSSION OF "A POSSIBLE RAINFALL PERIOD EQUAL TO ONE-NINTH THE SUN-SPOT PERIOD."

By DINSMORE ALTER.

[University of Kansas, Lawrence, Kans., Apr. 26, 1921.]

I have naturally been much interested in Prof. Marvin's conclusions[1] regarding my paper.[2] I am very sorry that it is impossible for us to agree concerning the possibility of the phenomenon discussed, and especially concerning the legitimacy of the method employed. A further statement concerning some of the points raised by him may be in order.

[1] MO. WEATHER REV., February, 1921, 49: 83–85.

[2] *Ibid.*, pp. 74–83.